View Classification of Color Doppler Echocardiography via Automatic Alignment between Doppler and B-mode Imaging

Jerome Charton[1*], Hui Ren[1*], Jay Khambhati[1], Jeena DeFrancesco[2], Justin Cheng[2], Anam A. Waheed[2], Sylwia Marciniak[2], Filipe Moura[2], Rhanderson Cardoso[2], Bruno B. Lima[2], Erik Steen[3], Eigil Samset[3], Michael H. Picard[1] Xiang Li[1], and Quanzheng Li[1]

[1] Massachusetts General Hospital
[2] Brigham and Women's Hospital
[3] GE Healthcare
*Joint first authors

Abstract. Echocardiography serves as a gold standard for diagnostic imaging in cardiovascular disease since it is non-intrusive, minimally invasive, and affordable. Recent advancements in deep learning techniques allowed the practical application of computer-assisted echocardiography imaging analysis, such as view classification, image segmentation, and disease diagnosis. However, unlike the more commonly investigated brightness (B-mode) imaging, there is limited research and open-source tools for the automatic processing of color Doppler echocardiography imaging (CDI) due to its more specific application and more heterogeneous image features (color flow overlaid on brightness images). Thus in this work, we developed a general framework to perform view classification of the Doppler echocardiography by leveraging the existing view classification algorithms (e.g., EchoCV) on B-mode imaging. Specifically, we developed a deep feature embedding-based module to automatically align CDI and B-mode videos based on the distance between their low-dimensional embedding. The proposed framework was evaluated on a dataset consisting of 250 subjects with ground-truth view labels by human annotators.

Keywords: Echocargiography · Doppler Imaging · View Classification

1 Introduction

Echocardiography (echo) has been widely used for the diagnosis of cardiac conditions thanks to its lowered cost, better portability, and non-invasive nature. In most of the imaging protocols for echocardiography, 2D videos from multiple cross-sectional views will be acquired by the sonographers, where each view of the imaging is determined by the position and orientation of the probe and illustrates a specific set of regions of the cardiac anatomy [1]. For example, apical (A2C, A3C, A4C, and A5C) views and parasternal long/short axis views are

generally recommended for evaluating regurgitation at the valves [2]. In the past decade, with the advancement of deep learning-based medical image analysis methods [3], various models for the view classification of echo imaging [4-6] have been developed, commonly by training a convolutional neural network to predict the view label of the input echo videos directly.

However, there are limited studies on the view classification of color Doppler echocardiography imaging (CDI), which has served as the key method for the diagnosis and quantification of valve regurgitation [7], ventricular hypertrophy [8], myocardial infarction [9], and interventional planning [10]. Color Doppler echocardiography characterizes and quantifies blood flow in the heart's chambers and valves, based on the movement of red blood cells relative to the transducer [11-13]. However, the lack of a publicly available dataset of CDI, such as the dataset presented in EchoNet for B-mode videos [14], makes it challenging to train and validate large-scale models. Nevertheless, computer-assisted automatic processing of CDI, especially the task of view classification, is crucial in both clinical practice (e.g., automated guidance for sonographers) and as a prerequisite step for the down-streamed tasks such as segmentation and risk assessment.

In response to the challenges above and the need for an accurate and robust view classification model for the CDI, while at the same time leveraging the existing view classification methods on B-mode videos, in this work, we developed an integrated view classification model for the CDI. The model will automatically align CDI and B-mode videos, then infer the view label of CDI based on the view label of the B-mode videos as predicted by EchoCV [4]. The model performed low-dimensional feature embedding on both B-mode videos and pre-processed CDI (with color flow removed) using EchoNet [14], then identified the link between B-mode/Doppler pairs based on their distance in the embedding space. The proposed model is trained and validated on a 250-subject dataset collected in this study, consisting of 2,189 color Doppler videos with manual-annotated labels for six different views (PLAX, PSAX, A2C, A3C, A4C, and A5C). The performance of the proposed model is evaluated based on the accuracy of view label prediction.

2 Methodology

2.1 Overview of the framework

As illustrated in Fig. 1, the proposed framework will take DICOM files as input. The pre-processing consists of an image series selection step to separate the CDI and B-mode videos. B-mode videos are then forwarded to an existing view classification module to generate the view labels. At the same time, both CDI and B-mode videos are sent to the image alignment module that will identify the correct video pairs. Finally, view labels estimated on the B-Mode videos are propagated to the CDI based on the pairing association.

Our proposed model integrates two modules in a single pipeline: the CDI/B-mode Alignment module by EchoNet [14] embedding and the View Classification

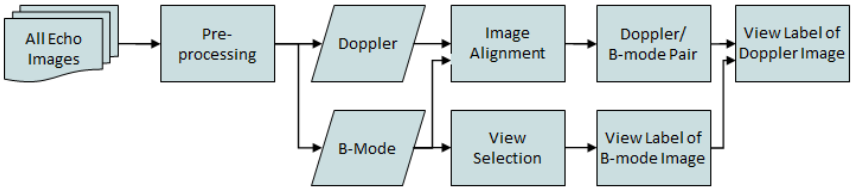


Fig. 1. Algorithmic pipeline of the proposed framework, consisting of the preprocessing module, B-mode view selection module (section 2.3), and image alignment module (section 2.4).

module by EchoCV [4]. EchoNet uses a Spatio-temporal convolution for its ejection fraction estimation. We have reused its R(2+1)D [15] network for embedding the CDI and B-Mode videos. EchoCV is an image-based deep learning model that can perform view classification, heart segmentation, measurement evaluation, and disease detection using VGG and U-Net networks. In our pipeline, we have used its VGG-13 view classifier network. This network takes a random number of frames from the input video, estimates the view label of each frame and averages the results. In the original EchoCV, as well as in our pipeline, ten randomly selected frames were used.

2.2 Data collection

This study used an in-house echocardiography dataset consisting of B-mode (without color) and color Dopplers (B-mode with colored blood flows burnt in) echocardiography videos from 250 subjects. These echocardiograph videos were acquired using various models and brands of machines, including 31,452 B-mode and 15,200 Dopplers from Philips and 176 B-modes and 66 Dopplers from GE's devices. Following the standard clinical protocol, the Doppler videos in this dataset would generally be accompanied by their corresponding B-mode counterparts, as sonographers usually rely on B-mode videos for the localization and view selection. We have asked a cardiologist to annotate 71 of these pairs for validation purposes. In addition, three sonographers performed view annotations on 2,201 Doppler videos and generated the view labels (PLAX, PSAX, A2C, A3C, A4C, or A5C) for each video.

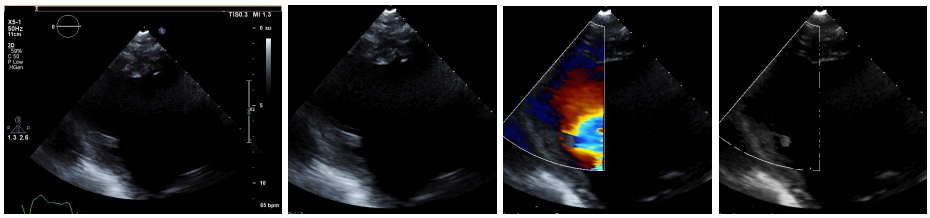
2.3 View classification on B-mode videos

EchoCV [4] was used for the view classification for B-modes videos without additional pre-processing and using ten randomly selected frames per video, following the protocol proposed in [14]. EchoCV proposed a view classification for 23 classes: PLAX remote, PLAX, PLAX zoom, PLAX centered, RV.inflow, PSAX APEX, PSAX PAP, PSAX MV, PSAX AoV, PSAX AoV zoom, A2C, A2C No occlusion, A2C occluded LA, A2C occluded LV, A3C No occlusion, A3C occluded LA, A3C occluded LV, A4C No occlusion, A4C occluded LA,

A4C occluded LV, A5C, Subcostal, Suprastemal, or OTHER. In contrast, our Doppler view labeling includes only PLAX, PASX, A2C, A3C, A4C, and A5C. Thus, We have merged sub-classes of PLAX from EchoCV into a single label of "PLAX", and pulled occluded LA and occluded LV sub-classes, PSAX APEX, PSAX PAP, and PSAX MV all together into the OTHER label.

2.4 Automatic alignment between Doppler and B-mode videos

For the CDI/B-mode alignment module, we have taken an extra preprocessing step to remove the color jets from the Doppler videos and replaced the region with black color (Fig. 2(d)), in order to improve the image-wise similarity between CDI and B-mode video. In addition, EchoNet proposed a "hard" cropping and masking (fixed margin sizes and fixed cone shape mask) of the data to remove the in-pixel meta-data (Fig. 2(a)) and preserve only the probe acquisition. However, this pre-processing was not suitable for every sample of our dataset. Therefore, we propose an adaptive filter for cropping and masking the meta-data. For every video, an activity map is calculated. This activity map measures the number of modifications of each pixel across the video. Based on the activity map and a threshold, pixels are distinguished into either background or foreground by low/high activity, respectively. The background pixels are then painted black, and black margins are automatically cropped to fit into a square shape. Finally, all 2D videos are resized to 112x112 as proposed by [14].



(a) Original (b) B-mode (c) Doppler (d) Doppler with color flow removed

Fig. 2. Pre-processing of the echocardiograms. (a) Original data with meta-information, (b) and (c) are the B-mode and the Doppler, respectively, after cropping and masking using our method, and (d) is (c) after color removal.

Preprocessed videos are then used as inputs for the ejection fraction prediction network of EchoNet. We obtained the outputs from the second last layer of the network for each video for embedding, which is a 512-D vector. Matchings between CDI and B-mode videos are then identified by locating the CDI/B-mode pairs with the maximized Cosine similarities between their embedding vector. In this way, we can find the corresponding B-mode videos for every CDI by leveraging their intrinsic relationship in the low-dimensional embedding space and

without the need for model training. Illustration of the embedding and alignment module can be found in Fig. 3

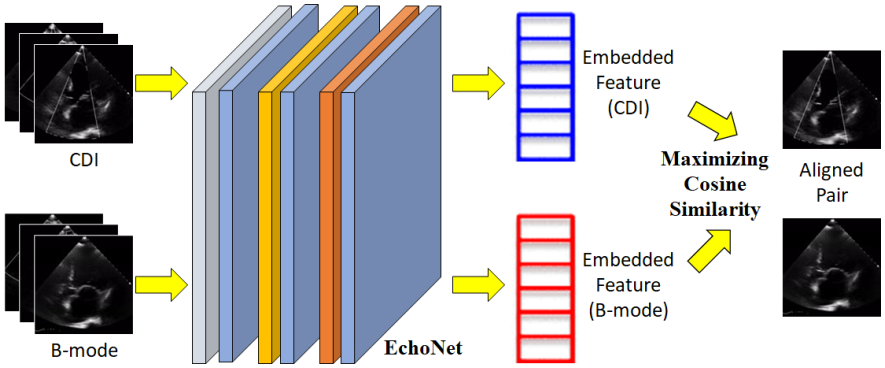


Fig. 3. Illustration of the image alignment module. Input CDI and B-mode videos are embedded by EchoNet into their corresponding 512-D feature vectors. CDI and B-mode videos are then aligned based on the cosine similarities between the embedded feature vectors.

3 Results and Discussion

	PLAX	PSAX	A4C	A5C	A3C	A2C	Other
PLAX	9	1	0	0	0	0	1
PSAX	1	5	0	0	0	0	1
A4C	1	1	20	0	0	0	1
A5C	0	0	1	5	0	0	0
A3C	0	0	0	0	18	1	1
A2C	0	0	0	0	0	4	0

Fig. 4. Confusion matrix of the predicted views by the proposed framework, on the sub-dataset with annotations of the ground-truth Doppler/B-mode pairing.

In order to validate the effectiveness of the B-mode/Doppler alignment module, we applied the proposed pipeline on the subset of data with human annotation for the ground truth pairing. As our proposed model relies on two consecutively coupled modules (EchoCV and the video alignment module) to work, the model error can be caused by either of the modules. Thus, we firstly used EchoCV to predict the view labels of the B-mode videos, then compared the prediction results to the manually-annotated labels of the CDI by the sonographers, based on the correspondence between CDI/B-mode videos as annotated by the cardiologists, in order to evaluate the effectiveness only for the EchoCV module. Results show that EchoCV can achieve a 6-view classification accuracy of 98.6%, with only one error case (misclassified of an A4C video to PLAX).

In the next step, we performed the same comparison between EchoCV outputs for the B-mode videos and manual annotations for the CDI, but based on the pair correspondence estimated by the alignment module. The B-mode/Doppler alignment result shows that 87.3% (62/71) of the pairs can be correctly aligned, while all the misaligned pairs ended up in misclassification of the view labels, as shown by the red highlighted values in the confusion matrix in Fig. 4. Combining both of the two modules, the proposed model can achieve the overall view classification accuracy of 85.9% in the 71-cases sub-dataset.

	PLAX	PSAX	A4C	A5C	A3C	A2C	Other
PLAX	591	24	11	1	5	1	31
PSAX	13	351	5	3	4	12	20
A4C	12	16	382	12	7	4	12
A5C	5	7	7	208	8	6	6
A3C	9	11	4	1	245	11	17
A2C	3	3	3	0	4	106	8

Fig. 5. Confusion matrix of the predicted views, on the whole dataset

In addition, we applied the proposed model to the full 250-subjects dataset, and the resulting confusion matrix is illustrated in Fig. 5. The overall classification accuracy on the full dataset is 86.0%, similar to the model performance on the 71-cases subset. Specifically, the model has achieved prediction accuracy of 89.0%, 86.0%, 85.8%, 84.2%, 82.2% and 83.5% for the views of PLAX, PSAX,

A4C, A5C, A3C, and A2C, respectively. With an average accuracy of above 85%, although further improvement is needed, our proposed model has the potential of helping novice sonographers to find the standard view not only in B-mode but also in color Doppler, alleviating the clinical work burden by directing the focus on the color Doppler views that of interest for further evaluation (e.g., valve regurgitation), avoiding human errors in view selection and assisting the imaging data quality check for research and management purposes.

4 Conclusion

In this work, we proposed a fully automatic and unsupervised model for the task of view classification on color Doppler imaging data. The model's performance validates that this task can be achieved by leveraging existing view classification tools (EchoCV) with an effective cross-modality (CDI/B-mode) alignment module without further need for model training or calibration. All the annotation labels used in this work are for model evaluation and validation, indicating the proposed model's robustness and applicability in practice.

With the promising results, we aim to test further the view classification performance of other similar models that were trained and working on B-mode videos. Also, by exploiting the relationship between CDI/B-mode videos, we aim to develop a framework that can more efficiently adapt other models on B-mode videos (e.g., LV segmentation) to CDI.

References

1. C. Mitchell et al., "Guidelines for Performing a Comprehensive Transthoracic Echocardiographic Examination in Adults: Recommendations from the American Society of Echocardiography," Journal of the American Society of Echocardiography, vol. 32, no. 1, pp. 1-64, 2019.
2. W. A. Zoghbi et al., "Recommendations for evaluation of the severity of native valvular regurgitation with two-dimensional and doppler echocardiography," Journal of the American Society of Echocardiography, vol. 16, no. 7, pp. 777-802, 2003.
3. J. H. Thrall et al., "Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success," Journal of the American College of Radiology, vol. 15, no. 3, pp. 504-508, 2018.
4. J. Zhang et al., "Fully Automated Echocardiogram Interpretation in Clinical Practice," Circulation, vol. 138, no. 16, pp. 1623-1635, 2018.
5. A. Madani, R. Arnaout, M. Mofrad, and R. Arnaout, "Fast and accurate view classification of echocardiograms using deep learning," NPJ digital medicine, vol. 1, no. 1, pp. 1-8, 2018.
6. A. Østvik, E. Smistad, S. A. Aase, B. O. Haugen, and L. Lovstakken, "Real-time standard view classification in transthoracic echocardiography using convolutional neural networks," Ultrasound in medicine & biology, vol. 45, no. 2, pp. 374-384, 2019.
7. C. M. Otto, The Practice of Clinical Echocardiography. Elsevier Health Sciences, 2007.
8. R. H. Anderson, E. J. Baker, A. Redington, M. L. Rigby, D. Penny, and G. Wernovsky, Paediatric cardiology. Elsevier Health Sciences, 2009.
9. A. Jeremias and D. L. Brown, Cardiac Intensive Care. Elsevier Health Sciences, 2010.
10. M. S. Norell, J. Perrins, B. Meier, and A. M. Lincoff, Essential Interventional Cardiology. Elsevier Health Sciences, 2008.
11. J. A. Jensen, Estimation of blood velocities using ultrasound: a signal processing approach. Cambridge University Press, 1996.
12. M. Meola, J. Ibeas, G. Lasalle, and I. Petrucci, "Basics for performing a high-quality color Doppler sonography of the vascular access," The Journal of Vascular Access, vol. 22, no. 1_suppl, pp. 18-31, 2021.
13. M. Bakircioglu, T. Sumanaweera, C. Bradley, P. Linyong, and J. Hossack, "Dynamic color Doppler extended field of view imaging," in 2001 IEEE Ultrasonics Symposium. Proceedings. An International Symposium (Cat. No.01CH37263), 2001, vol. 2, pp. 1423-1426 vol.2.
14. D. Ouyang et al., "Video-based AI for beat-to-beat assessment of cardiac function," Nature, vol. 580, no. 7802, pp. 252-256, 2020.
15. D. Tran et al. A closer look at spatiotemporal convolutions for action recognition. In Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition 6450–6459 (2018).